MAS364/61006: Assignment 2

Solutions and discussion are written in blue. Some common pitfalls are indicated in teal. A sample mark scheme is given in red, with each mark placed after the statement/deduction for which the mark would be given. As usual, mathematically correct solutions that follow a different method would be marked analogously.

Marks are given for [A]*ccuracy,* [J]*ustification, and* [M]*ethod.*

⋆ 1. (a) A random quantity θ is known to take real values, believed to be moderately close to zero. Arrange the following four candidate prior distributions for θ in order, from the most informative (first) to the least informative (last).

$$\mathrm{N}(0,5), \qquad \mathrm{N}(0,1), \qquad \mathrm{Cauchy}(0,1), \qquad \mathrm{Uniform}\left([-\tfrac{1}{4},\tfrac{1}{4}]\right)$$

(b) Which, if any, of the distributions in part (a) would usually be viewed as weakly informative?

Solution.

(a) $\mathrm{Uniform}\left([-\tfrac{1}{4},\tfrac{1}{4}]\right)$, $\mathrm{N}(0,1)$, $\mathrm{N}(0,5)$, $\mathrm{Cauchy}(0,1)$. [3A]

(b) $\mathrm{Cauchy}(0,1)$ is weakly informative [1A] (because it puts so much mass on values close to $\pm\infty$ that both its mean and variance fail to be defined).

The others are not weakly informative, [1A] for a parameter taking values in $\mathbb{R}$.

⋆⋆ 2. (a) The following statements and formulae are written in Bayesian shorthand. Write a version in precise mathematical notation, for continuous random variables.

i. $f(x,y)=f(x|y)f(y)$.

ii. If $(u,t)\sim \mathrm{NGamma}(m,p,a,b)$ then $u|t\sim \mathrm{N}(m,\frac{1}{pt})$.

(b) On the reference sheet you will find a conjugate pair involving both the Poisson and Gamma distributions. State this relationship using Bayesian shorthand.

Solution.

(a) i. $f_{(X,Y)}(x,y)=f_{X|_{\{Y=y\}}}(x)f_Y(y)$ [1A]

ii. If $(U,T)\sim \mathrm{NGamma}(m,p,a,b)$ then $U|_{\{T=t\}}\sim \mathrm{N}(m,\frac{1}{pt})$. [1A]

(b) If $x|\theta\sim \mathrm{Poisson}(\theta)^{\otimes n}$ and $\theta\sim\mathrm{Gamma}(\alpha,\beta)$ then $\theta|x\sim\mathrm{Gamma}\left(\alpha+\sum_{i=1}^n x_i,\,\beta+n\right)$. [2A]

⋆⋆ 3. (a) i. Show that the reference prior for the exponential distribution is given by

$$f(\lambda)\propto\begin{cases}\frac{1}{\lambda} & \text{for } \lambda>0\\ 0 & \text{otherwise.}\end{cases}$$

ii. Find the reference prior for the $\mathrm{N}(0,v)$ distribution.

iii. In each of i and ii, is the result a proper or improper prior density function?

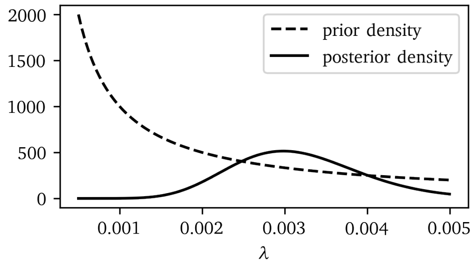
(b) Electronic devices are, in general, most likely to fail soon after they are manufactured or after their warranty period has expired. In between these times, their lifetime is often modelled using the Exponential distribution. A manufacturer of industrial high-temperature thermometers tracks a sample of twenty devices. The duration before the devices fail are measured and recorded in days as

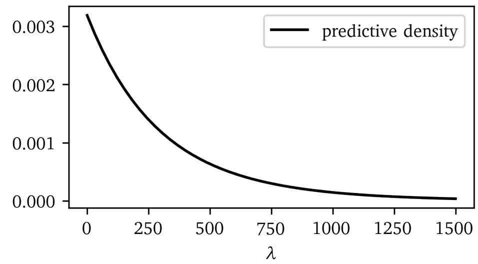
$$(3, 5, 9, 11, 23, 58, 89, 113, 124, 190, 221, 262, 279, 280, 281, 396, 565, 587, 647, 1062).$$

It is decided to to exclude devices that failed during their first 10 days from the analysis. With these removed, the resulting data $y_i = x_i - 10$ satisfies $\sum y_i = 5025$.

i. Using the reference prior from part (a) and the data y_i, find the posterior density of λ. Is this a proper or improper density function?

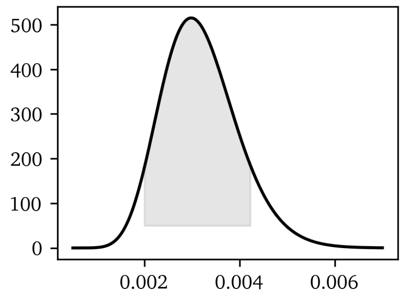
ii. A colleague produces the following plots of some of the density functions associated to the Bayesian update in part i. One of these plots contains an error. What is the error and what should be done to correct it?



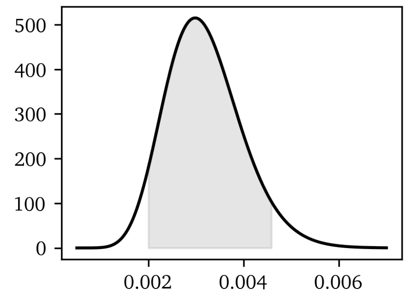


iii. It is required to give a high posterior density region for λ. Which one of the following plots best represents the results in this way? For each of the other plots, give a reason why you did not choose it.

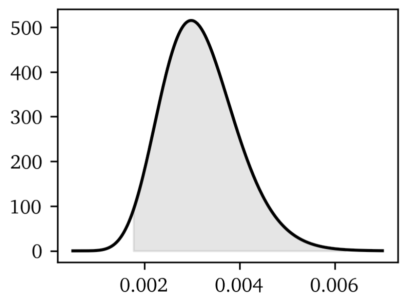
(A)



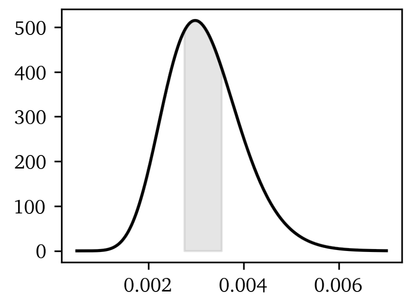
(B)



(C)



(D)



iv. A draft report of the analysis includes the sentence:

> *The lifetime of a thermometer was modelled by* Exp(λ)*, based on a reference prior and the available data.*

Is this a fair description of the analysis? If not, suggest a correction.

Solution.

(a) i. The likelihood function of $M_\lambda \stackrel{d}{=} \text{Exp}(\lambda)$ is

$$L_{M_\lambda}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x, \lambda \in (0,\infty) \\ 0 & \text{otherwise.} \end{cases}$$

[1M] On the non-zero part, we have $\log L_{M_\lambda}(x) = \log \lambda - \lambda x$ and hence (taking $X \sim M_\lambda$)

$$\mathbb{E}\left[-\frac{d^2}{d\lambda^2}\log L_{M_\lambda}(X)\right]^{1/2} = \mathbb{E}\left[\frac{1}{\lambda^2}\right]^{1/2} = \frac{1}{\lambda}.$$

[1M + 1A] Hence the reference prior has density

$$f(\lambda) \propto \begin{cases} \frac{1}{\lambda} & \text{for } \lambda \in (0,\infty) \\ 0 & \text{otherwise.} \end{cases}$$

ii. The likelihood function of $M_v \stackrel{d}{=} \text{N}(0,v)$ is

$$L_{M_v}(x) = \frac{1}{(2\pi v)^{1/2}} e^{-\frac{x^2}{2v}}$$

for $x \in \mathbb{R}$ and $v \in (0,\infty)$, and zero otherwise. [1M] On the non-zero part, $\log L_{M_v}(x) \propto -\frac{1}{2}\log v - \frac{x^2}{2}\frac{1}{v}$ and hence (taking $X \sim M_v$)

$$\begin{aligned}\mathbb{E}\left[-\frac{d^2}{d\lambda^2}\log L_{M_v}(X)\right]^{1/2} &\propto \mathbb{E}\left[\frac{1}{2v^2} + \frac{X^2}{2}\frac{2}{v^3}\right]^{1/2} \\ &\propto \left(\frac{1}{2v^2} + \frac{\mathbb{E}[X^2]}{v^3}\right)^{1/2} \\ &\propto \left(\frac{1}{2v^2} + \frac{v}{v^3}\right)^{1/2} \\ &\propto \frac{1}{v}\end{aligned}$$

[1M + 1A] Hence the reference prior has density

$$f(v) \propto \begin{cases} \frac{1}{v} & \text{for } \lambda \in (0,\infty) \\ 0 & \text{otherwise.} \end{cases}$$

iii. The reference prior is the same in parts i and ii (this is a coincidence!). It is easily checked that $\int_0^\infty \frac{1}{y}\,dy = \infty$, so these are improper priors. [1A]

(b) i. Using Baye's rule (extended to improper priors) [1M] we obtain that for $\lambda \in (0,\infty)$

$$f_{\Lambda|_{\{X=x\}}}(\lambda) \;\propto\; \left(\prod_{i=1}^n \lambda e^{-\lambda x_i}\right)\frac{1}{\lambda} \;\propto\; \lambda^{n-1} e^{-\lambda \sum_1^n x_i}$$

and zero otherwise. [2A] This is the density function of Gamma$(n, \sum_1^n x_i)$, and hence is a proper density function. [1A]

ii. The horizontal axis of the predictive density is labelled with λ. [1A] This is incorrect because the predictive density is for the quantity that we model (and not the parameter). A better axis label would be y. [1A]

iii. (A) is plotted incorrectly, the bottom of the HPD region is floating above the horizontal axis. [1J]
(B) is the best choice. [1A]
(C) is an interval of the form $[a, \infty)$, but we should (unless there is a specific reason otherwise) give an equally tailed region. [1J]
(D) is too thin to represent the range of likely values taken by λ. [1J]

iv. No – the quantity that we modelled is not stated correctly. We modelled the lifetime of a thermometer after its first ten days of usage (recall $y_i = x_i - 10$) conditional on the event that it does not fail during its first 10 days. [1A + 1J]

Total marks: 28